图书情报工作
第 65 卷 第 23 期 2021 年 12 月

科学数据开放共享合作关系实证研究*
——以北京大学开放研究数据平台为例

■ 张慧[1] 程雨琦[1] 王传清[2]
[1] 上海大学图书情报档案系 上海 200444 [2] 中国科学院文献情报中心 北京 100190

摘 要：[目的/意义]揭示北京大学开放研究数据平台 3 个代表性数据集的利用现状，为科学数据开放共享相关研究提供参考。[方法/过程]选择 CNKI 中利用 CFPS、CHARLS 和 CLHLS3 个数据集的论文作为研究对象，采用网络分析方法并借助相关工具，分别从作者、机构、主题维度对合作研究进行分析。[结果/结论]2011 年后各研究主体的年度合作度趋于稳定，随研究主体范围扩大，合作率逐渐降低。3 种中心度指标相互补充可发现重要机构。一级机构呈现以北京大学为核心，部分机构间保持合作，众多机构积极参与合作的情形；合作机构数目和总合作次数排名靠前的二级机构较稳定，相互间保持稳固合作关系，大致可将其分为 3 类机构。核心作者群表现出部分作者团体合作频繁且合作关系稳固，并存在中介作者可连通几个合作团体。合作研究论文涉及主题丰富，重点研究主题包括老年人身心健康、家庭消费及资产情况以及老龄化人口的社会保障。为促进科学数据开放共享的发展，未来可从建设权威性科学数据开放共享集成平台、融合学术数据库、组织学术竞赛等方面改进。

关键词：开放科学数据 CFPS CHARLS CLHLS 合作分析
分类号：G203
DOI：10.13266/j.issn.0252-3116.2021.23.003

1 引言

近年来，开放科学研究已引起全球科学界的广泛关注[1]，开放科学数据为其中的重要一环[2]。我国早在 2015 年 11 月 3 日公布的“十三五规划”中首次提出“实施国家大数据战略，推进数据资源开放共享”[3]，而后国务院办公厅颁布的《科学数据管理办法》确立了科学数据“开放为常态、不开放为例外”的共享原则[4]。开放科学数据在我国越来越得到高度重视，目前学界围绕科学数据开放共享中的数据治理[5-6]、影响因素[7]、共享政策[8-9]、机制模式[10-11]、利益相关者角色与责任[12-13]、共享平台[9]等主题进行了广泛探索。同时，国内外已建立众多科学数据开放共享平台[9]，包括我国正在建设中的 20 个国家科学数据中心[14-15]，和一些高校科学数据开放平台[16-17]。其中，北京大学开放研究数据平台就是一个建设典范。截至 2021 年 4 月 28 日，该平台已收录 73 个数据空间和 314 个数据集，拥有一批国内具有极高影响力的精品调查数据项目，如：中国家庭追踪调查（China family panel studies，CFPS）、中国健康与养老追踪调查（China health and retirement longitudinal study，CHARLS）、中国老年健康影响因素跟踪调查（Chinese longitudinal healthy longevity survey，CLHLS），现已更名为中国老年健康与家庭幸福调查（Chinese longitudinal healthy longevity survey and happy family，CLHLS-HF）等。这 3 个跟踪调查项目分别始于 2010 年、2011 年和 1998 年，均已成为具有规范性、规模性和周期性的调查数据集，并获广泛关注与利用[18]。此外，基于北京大学开放研究数据平台开展的“全国高校数据驱动创新研究大赛”，大力促进了科学数据的重复使用及学术产出。科学数据重用可有效避免科学数据重复采集、节省经费且提高效率，但同时也需耗费一定精力消化已有信息[19]。相较于传统科研合作关系，以科学数据重用为纽带的合作可积极推动科学数据开放共享的发展。因此，本文选择北京大学开放研究数据平台探索科学数据开放共享的合作关系。截至 2021 年 5 月已有 2 篇论文介绍了北京

* 本文系国家社会科学基金项目“开放科学环境下的科学数据开放共享机制与对策研究”（项目编号：18ATQ007）研究成果之一。
作者简介：张慧，讲师，博士，硕士生导师；程雨琦，硕士研究生；王传清，馆员，副编审，博士，通讯作者，E-mail：wangcq@mail.las.ac.cn。
收稿日期：2021-07-11 修回日期：2021-10-05 本文起止页码：22-30 本文责任编辑：徐健

大学开放研究数据平台建设[20－21]，但尚未发现此平台数据利用情况的研究。因此，本文基于北京大学开放研究数据平台的 CFPS、CHARLS、CLHLS3 个科学数据集，结合中国知网(CNKI)中利用这些数据发表的论文，采用网络分析方法和相关工具，拟对作者和机构合作研究做初步分析，以揭示北京大学开放研究数据的利用现状，并进一步促进科学数据开放共享的发展。

2 数据及研究方法

2.1 数据获取及处理

在 CNKI 专业检索界面进行检索，检索范围为学术期刊，分别输入检索式“SU% =(‘CFPS’+‘China Family Panel Studies’+‘中国家庭追踪调查’)”“SU% =(‘CHARLS’+‘China Health and Retirement Longitudinal Study’+‘中国健康与养老追踪调查’)”“SU% =(‘CLHLS’+‘Chinese Longitudinal Healthy Longevity Survey’+‘中国老年健康影响因素跟踪调查’+‘中国老年健康调查’+‘中国老人健康长寿影响因素跟踪调查’)”，并将检索起始日期分别限定为 2010 年 1 月 1 日、2011 年 1 月 1 日、1998 年 1 月 1 日，其他条件保持默认设定。截至检索日期 2021 年 4 月 28 日，得到期刊论文分别有 1 067、1 519 和 193 篇。

合并以上 3 类数据集相关期刊论文检索结果，且由于检索结果存在干扰因素，如主题词“CFPS”为“Cell-free protein synthesis”的缩写，或者“CHARLS”常出现在人名中，经过人工筛选并自编程序去重后最终得到期刊论文 1 493 篇，论文去重依据为若文章标题、期刊名称和出版年份一致，则为重复论文。因 2021 年数据不完全，故不纳入统计分析范畴。在北京大学开放科学数据利用研究中，2011 年前仅有 2 篇英文论文，但在 2011 年(含)之后，无论中文论文还是英文论文均呈现随年份显著增长的趋势，2018 年后增长速率大幅提升，年度发文量总数也呈现出相似趋势，此平台受到越来越多的关注与利用。

2.2 研究方法及工具

本文采用社会网络分析法，将合作度和合作率[22]作为分析北京大学开放研究数据合作研究的基本指标，并就不同研究主体分别从机构合作[23]和作者合作两个维度探讨合作情况，借助软件 Pajek 2.05 计算网络中心度[24]。涉及作者姓名去重、中英文对照、合并以及机构名称中英文对照和合并，均采用人工处理和自编代码辅助的方式。网络可视化则通过 VOSviewer 1.6.16 软件实现。

3 合作研究分析

3.1 数据预处理

为从机构合作和作者合作维度分析北京大学开放科学数据利用的合作研究，需基于 1 493 篇期刊论文对数据做进一步处理。

3.1.1 机构合作数据

本文机构合作分析含一级机构和二级机构两个层面(非高校仅涉及一级机构，如中国疾病预防控制中心)，其中高校一级机构为学校，如北京大学；二级机构为高校设置的学部或学院，如北京大学医学部、上海大学经济学院，机构数据均获取自各高校官网。同篇论文忽略机构合作次数，仅考虑合作关系。在 1 493 篇论文中，一级机构合作研究共 549 篇，占 36.77%，涉及 530 个一级机构；二级机构合作研究共 562 篇，占 37.64%，涉及 785 个二级机构。

3.1.2 作者合作数据

在 1 493 篇论文中，作者合作研究共 1 159 篇，占 77.63%，涉及 2 642 位作者。为更清晰地揭示作者利用此平台数据合作研究的情况，本文依据普莱斯定律筛选核心作者群[25]。发文量最多的作者为北京大学国家发展研究院赵耀辉教授，共发文 18 篇。核心作者群即选择发文数目 N > =0.749 $\sqrt{18}\approx 3$ 篇的作者，即本文核心作者群包含 199 位作者，后文则据此进行作者合作研究分析。

3.1.3 年度合作度和年度合作率

不同研究主体年度合作度和年度合作率变化见表 1，研究主体依次为一级机构、二级机构和作者。由表 1 可知，2011 年后各研究主体的年度合作度趋于稳定，

表 1 不同研究主体年度合作度、合作率变化统计

年份	一级机构		二级机构		作者	
	年度合作度/个	年度合作率/%	年度合作度/个	年度合作率/%	年度合作度/个	年度合作率/%
2009	3	100	4	100	4	100
2011	0	0	0	0	3	33
2012	2	43	2	57	3	71
2013	2	53	2	65	3	71
2014	3	48	3	54	3	76
2015	3	37	3	43	3	72
2016	3	37	3	44	3	81
2017	2	37	2	46	3	74
2018	2	33	3	39	3	79
2019	2	36	3	42	3	80
2020	3	35	3	43	3	78
2021	2	42	3	45	3	78

作者为 3 个,机构为 2 个或 3 个,表明每年利用此平台数据的合作研究中,平均每篇文章包含 3 位作者,涉及 2 个或 3 个一级或二级机构。此外,2011 年后各研究主体的年度合作率有明显区别与联系。区别体现在纵向比较,从左到右年度合作率依次约 30% –55% 、40% –65% 、70% –80% 。联系体现在横向比较,随研究主体范围扩大,合作率逐渐降低。

3.2 机构合作研究分析

本文从一级机构和二级机构两个层面分析中心度排名前 10 的机构利用此平台数据进行合作研究的情况。

3.2.1 机构中心度

合作网络中心度排名前 10 的机构及指标值如表 2 所示。点度中心度反映某机构合作机构总数。530 个一级机构中,北京大学与 106 个机构产生合作,点度中心度最大,遥遥领先;其次为浙江大学和杜克大学。785 个二级机构中,点度中心度最大的两个机构均来自北京大学,分别是经济与管理学部和医学部,但绝对数值相差较大,分别与 103 个和 39 个二级机构产生合作;其次为浙江大学医学院和中国疾病预防控制中心。

表 2 机构合作中心度排名前 10 机构及指标值一览

中心度	一级机构	值	二级机构	值
点度中心度	北京大学	106	北京大学经济与管理学部	103
	浙江大学	42	北京大学医学部	39
	杜克大学	36	浙江大学医学院	26
	中国人民大学	24	中国疾病预防控制中心	23
	山东大学	24	杜克大学医学院	19
	西南财经大学	22	昆山杜克大学全球健康研究中心	17
	中国疾病预防控制中心	21	武汉大学社会科学学部	16
	南加州大学	20	中国人民大学经济学部	16
	华中科技大学	17	北京大学社会科学学部	15
	密西根大学	16	西安交通大学医学部	15
接近中心度	北京大学	0. 394 3	北京大学经济与管理学部	0. 244 7
	浙江大学	0. 337 4	北京大学医学部	0. 202 3
	中国人民大学	0. 318 8	浙江大学医学院	0. 194 1
	杜克大学	0. 306 3	四川大学华西医院	0. 193 2
	山东大学	0. 303 8	杜克大学医学院	0. 189 8
	耶鲁大学	0. 303 8	西安交通大学医学部	0. 189 1
	南加州大学	0. 303 1	南加州大学经济学系	0. 188 8
	武汉大学	0. 302 1	北京大学理学部	0. 188 5
	中央财经大学	0. 301 9	河北医科大学附属第一医院	0. 187 6
	中南财经政法大学	0. 298 8	密西根大学医学院	0. 187 6
中介中心度	浙江大学	0. 004 5	北京大学经济与管理学部	0. 004 4
	南加州大学	0. 003 1	北京大学医学部	0. 001 7
	中国社会科学院	0. 002 6	浙江大学医学院	0. 000 7
	南京农业大学	0. 002 2	四川大学华西公共卫生学院	0. 000 6
	中国人民大学	0. 002 1	南京审计大学公共管理学院	0. 000 4
	杜克大学	0. 002 1	中国疾病预防控制中心	0. 000 4
	中国疾病预防控制中心	0. 001 2	四川大学华西医院	0. 000 4
	复旦大学	0. 001 0	山东大学公共卫生学院	0. 000 4
	华中科技大学	0. 001 0	南京大学社会学院	0. 000 3
	密西根大学	0. 000 9	西安交通大学医学部	0. 000 3

接近中心度反映网络中某机构与其他机构的接近程度。其排名前 10 的一级机构与点度中心度有 6 所机构相同,北京大学和浙江大学为网络中与其他机构最接近的两所高校。二级机构中,接近中心度排名前 5 的机构与点度中心度排名唯一不同之处在于四川大学华西医院代替中国疾病预防控制中心,并新增 5 所机构。

中介中心度反映某机构在网络中充当"桥梁"的程度。表 2 反映出一级机构排序前 10 中,此指标与前两个指标差异较大,且新增了中国社会科学院、南京农业大学和复旦大学 3 所机构。二级机构中北京大学经济与管理学部和医学部、浙江大学医学院稳居前 3,较前两个指标新增 4 个机构。

以上 3 个中心度均可反映网络中的重要机构,并

张慧，程雨琦，王传清. 科学数据开放共享合作关系实证研究——以北京大学开放研究数据平台为例[J]. 图书情报工作，2021,65(23):22－30.

可相互补充。综合以上结果，网络中排序前 10 的重要机构包括 17 个一级机构和 25 个二级机构，这些重要机构在下文分析中会进一步得以体现。

3.2.2 一级机构合作研究

图 1 为一级机构合作最大连通图，节点大小反映合作机构总数。一级机构共合作 549 篇论文，含 530 个一级机构，图 1 中含 447 个，表明 84.34% 的一级机构间可相互连通。显而易见，"北京大学"节点最大，表明其合作机构总数最多，与其他 106 个机构共合作 224 次，其中 11 个合作次数 > =5 次。106 个一级机构涉及 31 所国外高校、47 所国内高校及 28 个非高校机构。由此可见，北京大学的一级合作机构范围广泛，涉及国内、国外及非高校机构，且与部分高校之间保持着较强合作关系。

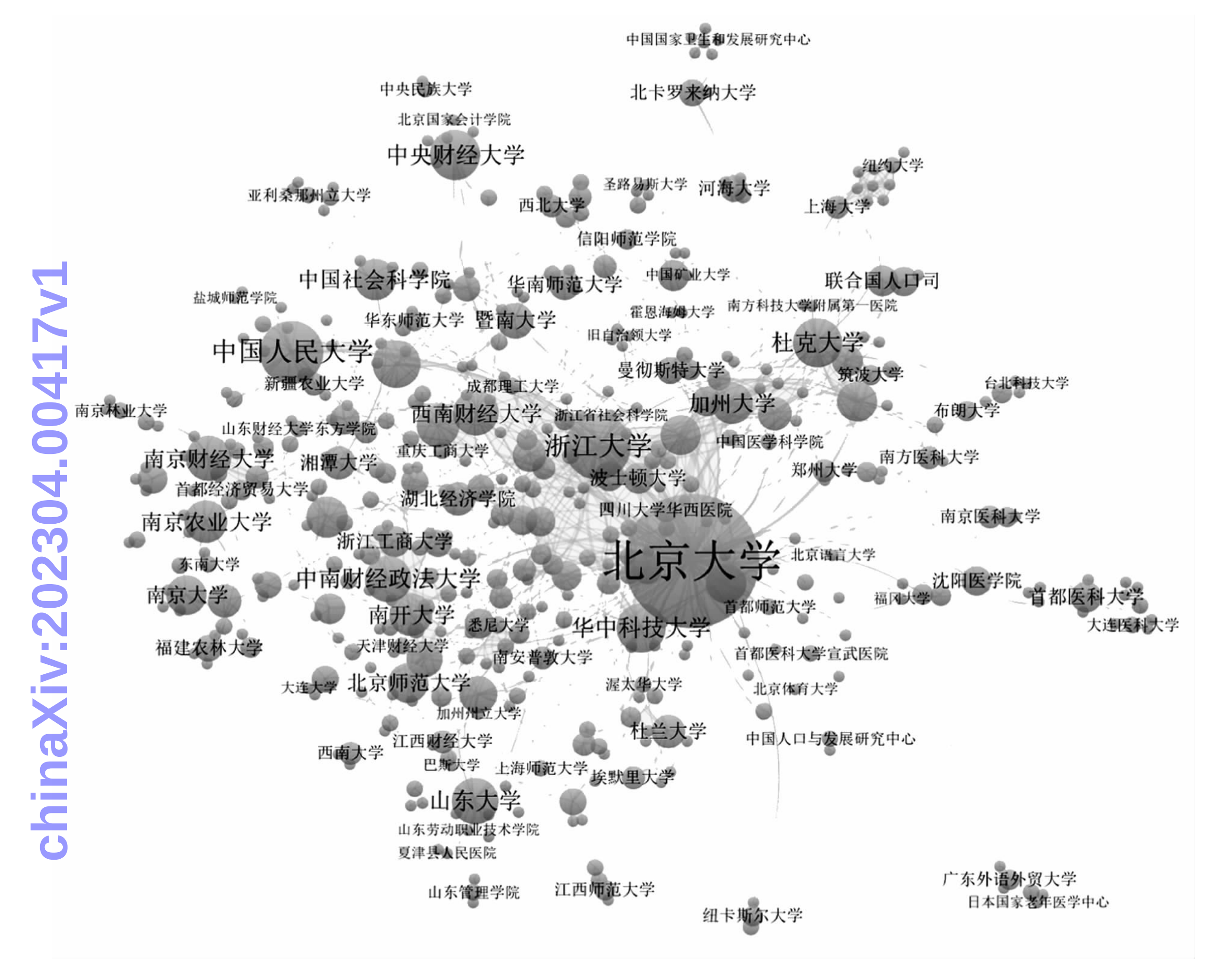


图 1 一级机构合作最大连通图(共 447 个节点)

合作机构总数排名第二为浙江大学，与 42 个一级机构合作了 61 次，除与北京大学合作 5 次外，与其他 6 个一级机构的合作次数 > =2 次；国外高校杜克大学与 36 个一级机构合作了 69 次，除与北京大学合作 8 次外，与其他 10 个机构的合作次数 > =2 次。530 个一级机构共产生 1 067 对机构合作，其中 909 对机构仅有 1 次合作，占 85.19%，表明大多数一级机构合作强度较低。整体来看，一级机构合作网络呈现以"北京大学"为核心，部分机构间保持合作，众多机构积极参与合作的情形。

3.2.3 二级机构合作研究

图 2 为二级机构合作最大连通图。二级机构合作研究共 562 篇，含 785 个机构，产生 1 239 次合作，最大连通图保留了 475 个节点，表明 60.51% 的机构间可相互连通。图中最大节点为北京大学经济与管理学部，其合作机构数目最多，与其他 103 个机构合作了 170 次，其中与之合作次数排名前 3 的机构依次为杜克大学医学院(7 次)、北京大学医学部(6 次)以及中国疾病预防控制中心(6 次)；北京大学医学部的合作机构总数排名第二，共与 39 个机构合作了 59 次；浙江大学医学院与 26 个机构产生了 30 次合作，节点大小排名第三，总合作次数排名第五；中国疾病预防控制中心与 23 个机构产生了 41 次合作，节点大小排名第四，总合



图2 二级机构合作研究最大连通图(共475个节点)

作次数排名第三;杜克大学医学院与19个机构产生了35次合作,节点大小排名第五,总合作次数排名第四。

二级机构合作网络中合作机构数目和总合作次数排名靠前的机构较稳定,且相互间保持稳固合作关系。整体来看,1 239对二级机构合作中1 100对机构仅产生1次合作,占88.78%;仅有9对机构合作次数>=5次,表明大部分二级机构的合作关系没有保持,仅少数机构维持着较为稳固的合作关系。此外,参与合作的二级机构大致可分为3类,一类为医学类部/院(含公共卫生学院),二类为经济管理类部/院,三类为社会学类部/院。此机构分布与本文研究数据集有关,CFPS数据集多用于经济类或社会学类主题研究,CHARLS和CLHLS数据集多用于医学类相关主题研究。

3.3 作者合作研究分析

由199位作者形成的"核心作者群"合作网络如图3所示,节点大小反映某作者的合作总次数。图3存在9个较明显的合作网络(见①-⑨),其中①号、②号和⑦号网络增加小图展示用于后文分析。①号网络含15位作者,人数最多。依据原始数据可知,此网络中有4位作者的合作总次数位列总网络前五,分别为中国疾病预防控制中心施小明(Xiaoming Shi),共发表11篇论文,与40位作者产生90次合作;北京大学健康老龄与发展研究中心曾毅(Yi Zeng),与45位作者合作发表10篇论文,产生88次合作;中国疾病预防控制中心殷召雪(Zhao-Xue Yin)和吕跃斌(Yuebin Lv)均发表7篇论文,分别与23位和31位作者均产生65次合作;此外,施小明和曾毅互为最多合作次数作者,殷召雪、吕跃斌的最多合作次数作者分别为施小明和曾毅、施小明,网络中其余作者的合作总次数均>20次。由此可见,此网络中作者合作表现较为活跃,且最活跃的作者间保持着稳固合作关系。

②号和⑦号网络的合作人数>=10位。②号网络含14位作者,其中来自北京大学国家发展研究院的赵耀辉共与39位作者产生86次合作;与之合作的美国南加州大学经济学教授J. Strauss与28位作者合作了62次,且互为最多合作次数作者。⑦号网络含11位作者,其中来自首都医科大学卫生管理与教育学院的景汇泉与34位作者合作了44次,来自北京大学中国社会科学调查中心丁华与30位作者合作了38次,两位为此网络总合作次数最多的作者,且相互合作次

图3 “核心作者群”合作网络



数最多。其余6个网络合作人数为6－8个，①－⑥号网络为连通网络（见图中胶囊形框）。计算作者中介中心度可知，曾毅、施小明、景汇泉、丁华为排名前列的中介作者。综上，核心作者合作网络表现出部分作者团体合作频繁且合作关系稳固，并存在中介作者可连通几个合作团体的情况。

3.4 重点研究主题的作者合作分析

对作者合作研究的1 159篇论文关键词进行中英文对照、同义词合并、无意义词（如CFPS）去除处理后得到2 062个关键词，为清晰揭示重点研究主题，选择出现频次＞＝9次的关键词网络（见图4）。VOSviewer将71个关键词自动聚为7类，结合具体研究内容可提炼为以下3类重点研究主题：

3.4.1 老年人身心健康

涉及关键词如慢性病、糖尿病、高血压、抑郁症状、精神健康、认知障碍、生活满意度、幸福感等。此主题下有3个作者团体合作发文较多，团体1以Changwei Li为代表，共围绕此主题发表8篇论文，涉及21位合作作者，其中与Tingting Liu和Zhiyong Zou分别合作6次和5次，与Jun Ma等7位作者合作2次，与其余12位作者仅合作1次，研究聚焦于中老年人的身体健康问题，如高血压、关节炎、糖尿病等；团体2以曾毅为代表，共参与发表7篇论文，与38位作者产生合作，其中与施小明和殷召雪分别合作6次和5次，与吕跃斌等11位作者合作2次及以上，与其余25位作者仅有1次合作，研究侧重老年人的认知功能；团体3以顾大男为代表，在其发表的6篇论文中与10位作者产生合作，其中与7位作者的合作次数在2次及以上，研究关注影响老年人健康的医疗保健问题。

3.4.2 家庭消费及资产情况

主要关键词如收入分配、消费结构、商业保险、土地流转，其中还涉及多维贫困问题，相关关键词如精准扶贫、减贫效果、灾难性卫生支出、收入差异、城乡差异等。此主题中未发现明显的作者合作团体，少部分作者合作发表了大于1篇的论文，如王晓全围绕“家庭商业保险”发表了3篇论文，涉及4位合作作者，除与贾

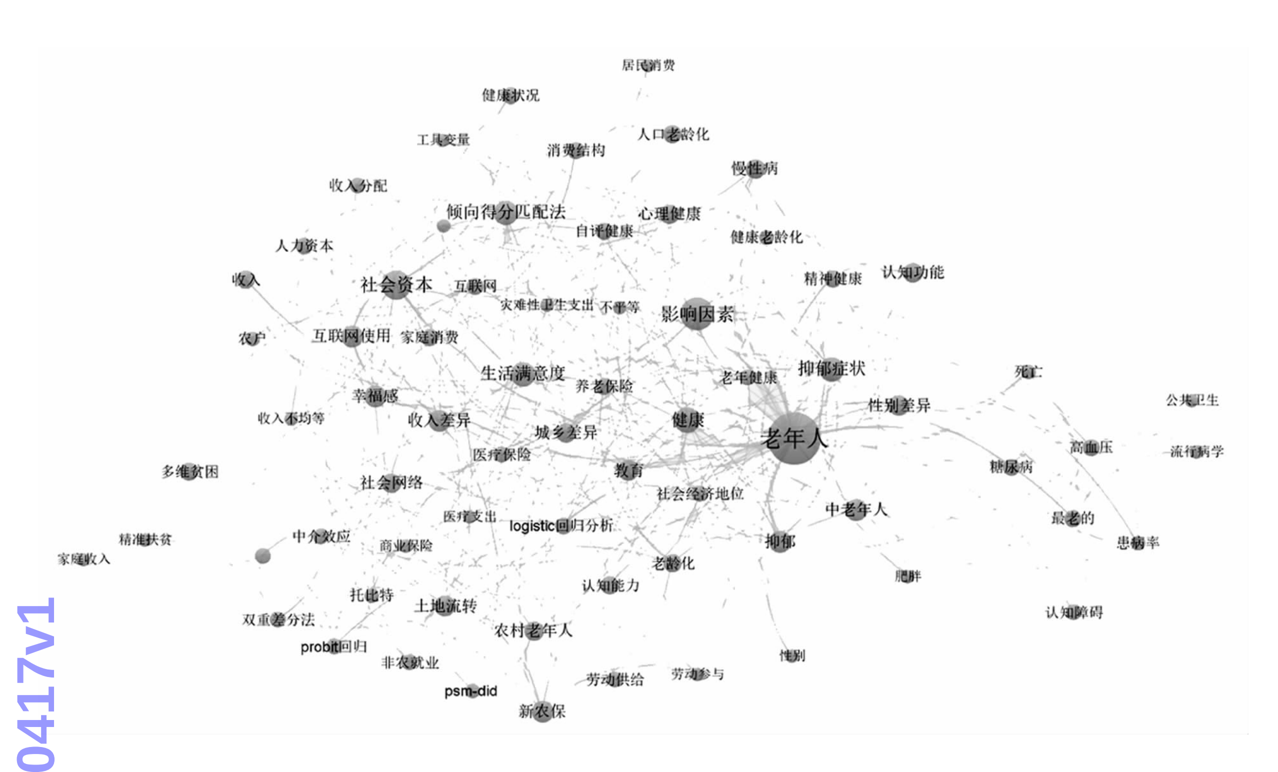


图 4 出现频次 > =9 的关键词聚类(共 71 个节点)

吴文有 2 次合作外,其余作者仅产生 1 次合作;张奇林共与 4 位作者合作发表了 3 篇关于"家庭贫困"的论文,仅与刘世爱产生了大于 1 次的合作;钱龙发表了 3 篇有关"农户土地流转"的论文,涉及 4 位合作作者,但均只产生一次合作。

3.4.3 老龄化人口的社会保障

涉及关键词如新农保、医疗保险、养老保险、社会经济地位等。此主题下李建新与夏翠翠之间保持着较为稳固的合作关系,合作发表 3 篇论文。其余大部分作者仅产生 1 次合作关系,如关于"新农保"研究的 18 篇论文中,除王小增和王林萍产生 2 次合作外,其余 39 位作者之间仅产生一次合作。统计各主题作者发文情况可知,多数作者仅发表过 1 篇论文,发文 >1 篇平均占比仅为 11.52% 。

综上所述,作者合作研究涉及主题丰富,研究群体重点关注"家庭"和"老年人",家庭方面重点探讨金融资产及消费,老年人群体着重研究身心健康及生活保障。此外,重点研究主题下绝大多数作者仅有 1 次发文,作者合作分析反映出仅少数发文较活跃的作者可与部分合作者保持合作,但只产生 1 次合作的现象更为普遍。

4 结论与讨论

通过上述实证研究,本文有如下 3 个主要结论:①基于北京大学开放研究数据平台 3 个数据集的现有研究在 2011 年后年度合作度趋于稳定,平均每篇文章含 3 位作者,涉及 2 个或 3 个一级或二级机构。随研究主体范围扩大,合作率逐渐降低。3 种中心度指标均反映出网络中的重要机构,并可相互补充。中心度排名前 10 的重要机构包括 17 个一级机构和 25 个二级机构。②一级机构合作网络呈现以"北京大学"为核心,部分机构间保持合作,众多机构积极参与合作的情形。二级机构合作网络中合作机构数目和总合作次数排名靠前的机构较稳定,且相互保持稳固合作关系。参与合作的二级机构大致可分为 3 类,一类为医学类部/院,二类为经济管理类部/院,三类为社会学类部/院。核心作者合作网络表现出部分作者团体合作频繁且合作关系稳固,并存在中介作者可连通几个合作团体的情形。③作者合作研究涉及主题丰富,重点研究主题包括老年人身心健康、家庭消费及资产情况,以及老龄化人口的社会保障。研究群体重点关注"家庭"和"老年人"。此外,重点研究主题下绝大多数作者仅参与 1 次发文,作者合作分析反映出仅有少数发文较活跃的作者可与部分合作者保持合作。

基于北京大学开放研究数据平台 3 个数据集对科学数据开放共享合作关系的实证研究说明 3 个问题:①无论是机构还是作者层面,对现有开放研究数据

的合作利用范围较为局限，参与机构以北京大学及其学部为主，参与活跃作者中曾毅、赵耀辉等均来自北京大学。②各研究主体的合作关系绝大多数未继续维持，超过85%的机构间和作者间仅有一次合作。③合作研究主题依赖于开放科学数据的类型和内容，即以科学数据重用为纽带的合作关系依赖于科学数据本身，一定程度上限制了此种形式的合作维系。

为促进科学数据开放共享的发展，本文就以上问题提出如下3点建议：①统一开放科学数据获取来源，建设权威性科学数据开放共享集成平台。目前国内已有平台各有侧重，如国家科学数据中心以科技数据为主，本文使用平台则以社科类数据为主，而权威性集成平台的建设有助于保证科学数据的全面性和统一性，进一步扩大用户范围，可缓解本研究反映的参与机构、作者范围和合作研究主题局限等问题。同时，基于此平台产生的科研成果可作为更全面的实证分析素材。②科学数据开放共享集成平台建设可融合学术数据库，将相关科研成果及研究主题、发文活跃作者等信息纳入科学数据开放共享集成平台建设中，为用户提供便捷式服务。③基于集成平台组织大范围的科学数据竞赛等学术性活动，积极推动科学数据的重复使用及高效利用。

本文仅利用北京大学开放研究数据平台3个数据集的现有研究来分析科学数据开放共享的合作研究有一定的局限性，今后可利用北京大学开放研究数据平台的所有数据集以及其他开放科学数据集在更大规模上开展进一步研究。

参考文献：

[1] UNESCO. UNESCO mobilizes 122 countries to promote open scienceand reinforced cooperation in the face of COVID-19[EB/OL]. [2021-11-19]. https://en.unesco.org/news/unesco-mobilizes-122-countries-promote-open-science-and-reinforced-cooperation-face-covid-19.
[2] 黄如花，赵洋，黄雨婷. 国际开放科学研究进展[J]. 图书情报工作，2021,65(1):140-149.
[3] 人民网. 中共中央关于制定十三五年规划的建议(全文)[EB/OL]. [2021-11-19]. http://politics.people.com.cn/n/2015/1103/c1001-27772701.html.
[4] 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2021-11-19]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.
[5] 盛小平，郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作，2020,64(22):25-36.
[6] 盛小平，田婧，向桂林. 科学数据开放共享中的数据质量治理研究[J]. 图书情报工作，2020, 64(22):11-24.
[7] 盛小平，袁园. 国内外科学数据开放共享影响因素研究综述[J]. 情报理论与实践，2021,44(8):173-179.
[8] 温芳芳. 国外科学数据开放共享政策研究[J]. 图书馆学研究，2017(9):91-100.
[9] 盛小平，武彤. 国内外科学数据开放共享研究综述[J]. 图书情报工作，2019,63(17):6-14.
[10] 李成赞，张丽丽，侯艳飞，等. 科学大数据开放共享：模式与机制[J]. 情报理论与实践，2017, 40(11):45-51.
[11] KITCHIN R, COLLINS S, FROST D. Funding models for open access digital data repositories[J]. Online information review, 2015, 39(5):664-681.
[12] WESSELS B, FINN R L, LINDE P, et al. Issues in the development of open access to research data[J]. Prometheus, 2014, 32(1):49-66.
[13] 盛小平，吴红. 科学数据开放共享活动中不同利益相关者动力分析[J]. 图书情报工作，2019, 63(17):40-50.
[14] 张先恩. 国家科学数据共享工程[J]. 科学中国人，2004(9):11-13.
[15] 国家科技基础条件平台中心. 国家科学数据中心[EB/OL]. [2021-11-19]. https://www.escience.org.cn/data-center/.
[16] 中国人民大学中国调查与数据中心. 中国综合社会调查[EB/OL]. [2021-11-19]. http://cgss.ruc.edu.cn/index.htm.
[17] 中国人民大学中国调查与数据中心. 中国老年社会追踪调查项目概况[EB/OL]. [2021-11-19]. http://class.ruc.edu.cn/xmjs/xmgk.htm.
[18] 北京大学开放研究数据平台. 北京大学开放研究数据平台简介[EB/OL]. [2021-11-19]. https://opendata.pku.edu.cn/about.xhtml;jsessionid=45d5426f10e05399-8f4f63bcf98b.
[19] 李佳潞. 科研人员数据重用行为影响因素及促进策略研究[D]. 长春：东北师范大学，2019.
[20] 罗鹏程，朱玲，崔海媛，等. 基于Dataverse的北京大学开放研究数据平台建设[J]. 图书情报工作，2016,60(3):52-58.
[21] 朱玲，聂华，崔海媛，等. 北京大学开放研究数据平台建设：探索与实践[J]. 图书情报工作，2016,60(4):44-51.
[22] 赵蓉英，余波. Altmetrics Top100论文的合作模式及其影响力因素分析[J]. 情报科学，2020, 38(4):3-8.
[23] 魏瑞斌. 科学计量学领域科研机构合作网络演化分析[J]. 情报杂志，2012,31(12):40-45.
[24] 张雪，张志强，陈秀娟. 基于期刊论文的作者合作特征及其对科研产出的影响——以国际医学信息学领域高产作者为例[J]. 情报学报，2019,38(1):29-37.
[25] 黄丽霞，纪苏桐. 基于SNA国内阅读推广领域作者合作关系研究[J]. 图书情报工作，64(7):119-126.

作者贡献说明:
张慧:负责论文选题、撰写与修改;
程雨琦:协助数据处理;
王传清:负责论文选题与修改。

An Empirical Study on Collaborative Relation of Scientific Data Open Sharing: Based on Peking University Open Research Data Platform

Zhang Hui[1] Cheng Yuqi[1] Wang Chuanqing[2]
[1] School of Library, Information and Archives, Shanghai University, Shanghai 200444
[2] Natioanl Science Library, Chinese Academy of Science, Beijing 100190

Abstract: [**Purpose/significance**] This paper aims to reveal the utilization status of the three representative datasets in Peking University Open Research Data Platform and provide references for the related research of scientific data open sharing. [**Method/process**] This paper selected CNKI papers using CFPS, CHARLS and CLHLS as the research objects, and adopted network analysis method and relevant tools to analyze their collaborative research from the dimensions of authors, institutions and themes. [**Result/conclusion**] After 2011, the annual collaboration degree of the research subjects has tended to be stable, and the collaboration rate gradually decreased with the expansion of the research subjects. The three centrality indicators can complement each other to find important institutions. The first-level institutions show that Peking University is the core, some institutions maintain collaboration, and many institutions actively participate in collaboration. The top-ranked second-level institutions in terms of the number of collaborative institutions and the total number of collaboration times are relatively stable, and maintain a solid collaborative relationship mutually, which can be roughly classified into three categories. The core author groups show that some author groups collaborate frequently and stably, and there are intermediary authors connecting to several collaborative groups. Moreover, collaborative research papers cover a variety of topics, focusing on physical and mental health of elderly people, family consumption and assets, as well as social security issues for the aging population. It can be improved in terms of building an authoritative integration platform of scientific data open sharing, integrating academic databases, and organizing academic competitions to promote the development of scientific data open sharing in the future.

Keywords: open scientific data CFPS CHARLS CLHLS collaboration analysis